REMARKS

Claims 1 through 12 are pending in this application.

Claims 1-12 were rejected.

Claim 1 and 6 have been amended in this Response.

In the following, the Examiner's comments are included in bold, indented type, followed by the Applicant's remarks:

> **Claims 1-3 and 6-12 are rejected under 35 U.S.C. 112, first paragraph, as failing to comply with the enablement requirement. The claim(s) contains subject matter which was not described in the specification in such a way as to enable one skilled in the art to which it pertains, or with which it is most nearly connected, to make and/or use the invention. The variables R, $T_{1...R}$, T_S, T_V, T_X, v, x, s, I, T_U, u, T_{UY} and Y set forth in the claims are not defined in the specification to enable one having ordinary skill in the art to understand the use the invention as claimed. Applicant is advised to amend the claims by defining the variables set forth in the claims. Applicant is reminded that no new matter should be added.**

Applicants have amended claims 1 and 6 based on the Examiner's advice. The claim amendments are supported by the disclosure as filed. The claimed subject matter is generally described in the specification as originally filed on page 6, line 19 through page 8, line 15, which includes pseudocode (page 7, line 23 through page 8, line 12), and in Figs. 4, 5 and 6. An example is provided on page 8, lines 24-32 and Figs. 7, 8 and 9. An example SQL implementation flow for the string clustering technique is provided in Table 1 on page 9. The amended claims are further supported by the claims as originally filed, as understood by a person of ordinary skill in the art. Applicants respectfully request that the Examiner withdrawal the rejects of claims 1-3 and 6-12.

> 6. **Claims 4 and 5 rejected under 35 U.S.C. 102(e) as being anticipated by Chandrasekar et al., (hereinafter "Chandrasekar") US Patent no. 6,578,032.**
>
> **As to claim 4, Applicant should duly note that a N-gram is a string of characters that may comprise all or part of a word. In particular, Chandrasekar the claimed "identifying unique n-grams in each string" col. 2, lines 6-7; "associating each string with zero or more cluster associated with a low frequency n-grams from that string" (col. 2, lines 10-20, lines 25-37); and "associating each string with zero or more clusters associated**

with low-frequency pairs of high frequency n-grams from that string" (col. 11, lines 59-67; col. 12, lines 1-12).

Applicants respectfully disagree. Chandrasekar fails to teach each element of claim 4. Claim 4 requires, in part, "associating each string with zero or more clusters associated with low frequency n-grams from that string." The portions of Chandrasekar cited for this limitation states:

> If the commonality between the text string and the existing cluster members satisfies a pre-defined threshold, the text string is added to the cluster. **If, on the other hand, the commonality does not satisfy the pre-defined threshold, a new cluster may be created. Each cluster may have a selected topic name.**
>
> . . .
>
> Each character string comprises a word or a phrase. The method comprises the steps of receiving at least one character string, and clustering a first character string with another character string into one or more groups, when the first character string satisfies a predetermined degree of commonality with one or more character strings in each of these groups. **When the first character string does not satisfy the predetermined level of commonality with another character string, another group is created.** The method also selects at least one of the character strings in each of the groups to be the group's topic name. Selection of the topic may be based on a pre-designation or a frequency of the received character strings with the groups. The selected topic may then be outputted.
>
> . . .
>
> In step 1004, QCluster Program 305 may calculate the frequency of the occurrence of the individual words and whole query. In step 1005, the **highest frequency** words and queries are determined, based on step 1004. The precise number of selected **highest frequency** "items" (i.e., words and/or queries) may vary, depending on the relative scores. For example, the two **highest frequency** items may be selected when their frequency scores are relatively close. On the other hand, only one **highest frequency** item may be selected, where the subject item has a frequency score that is **significantly higher** than the second highest frequency item. If two or more **highest frequency** items are selected, it is determined whether the items have the same frequency score, in step 1006. If the scores are not the same, the **highest frequency**

> item may be selected as the topic. Alternatively, a predetermined number of **highest frequency** items may be selected to be the topics. If the **highest frequency** items have the same frequency score, a predetermined criterion may be used to break the tie, in step 1008. For example, it may be that the longest item (i.e., the item with the most characters) is selected as the topic.

Chandrasekar, 2:10-20, 2:25-37, 11:59-67, 12:1-12 (emphasis added).

The emphasized portions of the cited passage teach away from "associating each string with zero or more clusters associated with **low frequency** n-grams from that string," and "associating each string with zero or more clusters associated with **low-frequency** pairs of high frequency n-grams from that string" as required by claim 4. In particular, the portion of Chandrasekar in col. 11, line 59-col 12, line 14 generally discusses "calculating[ing] the frequency of the occurrence of the individual words and whole query" and determining "the **highest frequency** words and queries." (emphasis added) Indeed, while the words "highest frequency" are used throughout the cited passages, as indicated above, the phrase used in claim 4, "low frequency," which has a clearly different meaning, is nowhere to be found. Thus, the cited portion of Chandrasekar teaches away from that which is claimed in claim 4 and the rejection of claim 4 should be withdrawn.
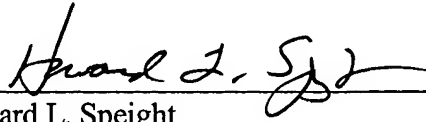
> As to claim 5, Chandrasekar the claimed "where a string does not include any low-frequency pairs of high frequency n-grams associating that string with clusters associated with triples of n-grams including the pair" (col. 11, lines 59-col. 12, line 12; col. 10, line 18-65).

The Examiner's explanation of the rejection of claim 5 is not understood. In any case, because dependent claims 5 includes all of the limitations of claim 4, which Applicants have shown to be patentable, Applicants respectfully request that the Examiner withdraw the rejection of claims 5.

SUMMARY

Applicants contend that the claims are in condition for allowance, which action is requested. Applicants do not believe any fees are necessary with the submitting of this response. Should any fees be required, Applicants request that the fees be debited from deposit account number 50-1673.

Respectfully submitted,

Howard L. Speight
Reg. No. 37,733
Baker Botts L.L.P.
910 Louisiana
Houston, Texas 77002
Telephone: (713) 229-2057
Facsimile: (713) 229-2757
E.Mail: Howard.Speight@bakerbotts.com

Date: December 20, 2005
ATTORNEY FOR APPLICANTS